DOCUMENT RESUME

ED 338 697 TM 017 533

AUTHOR McDonnell, Lorraine M.; Ormseth, Tor
TITLE Developing Indicators of Student Coursework.
INSTITUTION Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; Rand Corp., Santa Monica, Calif.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO CSE-TR-301
PUB DATE Mar 89
CONTRACT G0086-003
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Courses; *Curriculum Evaluation; Curriculum Research; *Educational Assessment; Educational Policy; Educational Research; Effective Schools Research; Elementary Secondary Education; *Program Design; *Research Design; Research Projects
IDENTIFIERS *Educational Indicators; *School Reform Assessment Project

ABSTRACT

The School Reform Assessment (SRA) project began in 1987 with the objective of developing indicators of student coursework that reliably and validly measure this central feature of schooling, while remaining sensitive to major policy changes and the information needs of policymakers, and efficiently collecting and reporting data. This paper describes the major research and policy positions that shaped the SRA project and then outlines the effort to meet those research and policy requirements in the study design. Because data collection was not complete, results of the study are not reported, but some of the lessons of indicator development learned along the way are discussed. The first section of the paper summarizes the major dimensions that curriculum research has identified as important in the development of indicators. The second section examines features of policy-relevant indicators. The third describes the current status of coursework indicators. A final section outlines the SRA project design. A 22-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED338697

C·R·E·S·S·T

Center for Research
on Evaluation, Standards,
and Student Testing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

DEVELOPING INDICATORS OF STUDENT COURSEWORK

CSE Technical Report 301

Lorraine M. McDonnell
Tor Ormseth

The RAND Corporation

TM017533

BEST COPY AVAILABLE

The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by this agency should be inferred.

This paper was written as part of a research project sponsored by the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST). It was prepared for presentation at the Annual Meeting of the American Educational Research Association, San Francisco, March, 1989.

Please address inquiries to: CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, California, 90024-1521

Introduction

This paper chronicles a tale of attempting to serve two masters: The professional cannon that educational indicators be reliable and valid, and the political imperative that they be useful and easily collected.[1] As statistics that reflect important characteristics of the educational system, indicators must accurately measure those features and adequately support the inferences that may be drawn from them. A deep concern about reliability and validity is especially critical in developing curriculum indicators—the focus of the School Reform Assessment project—because of the complexity of the phenomenon and the potential that measures may be overly simplified and thus mask significant variation across schools, classrooms, and students.

Yet over the past few years, indicator design has become as much a visible political enterprise as a technical task. Traditional concerns about reliability and validity have been joined by an interest in developing measures that are easily understood by policymakers and the lay public; that can be used to inform policy as well as practice; and that can be efficiently collected and reported. A growing demand for greater accountability in public education has prompted this increased attention to the development and use of educational indicators. Such a focus is predicated on the assumption that if sufficient data are available about schools, that information will serve as a resource for policymakers, concerned professionals, and the public, all of whom can use it to demand or effect improvements in schooling. As a result of this policy focus, educational indicator data are being used not just to describe the status of public education, but also as a basis for policy action. As of 1987, half of the 50 states use their indicator systems to trigger substantial policy actions that either reward, punish, or assist schools (OERI State Accountability Study Group, 1988).

It was in this environment that the School Reform Assessment (SRA) project began, late in 1987. The project, a joint effort of the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST) and The RAND Corporation, and funded by the U.S. Department of Education, has as its purpose developing indicators of student coursework[2] that reliably and validly measure this central feature of schooling, and that also: (a) are sensitive to major policy changes, (b) address policymakers' information needs, and (c) are efficient to collect and report.

We took the position that the demand for educational indicators to serve policy needs was not likely to disappear (despite academic misgivings), and that our responsibility as researchers was to assist in designing indicators that also met cannons of good social science. We believed that the history of earlier efforts held an important lesson for researchers. The social indicators movement of the 1960s failed to "deal with the style, objectives and constraints of decision-makers" (de Neufville, 1975). As a result, technical quality was not sufficient to guarantee the continuation of those efforts (de Neufville, 1975; MacRae, 1985): Indicator systems must also produce information useful to the policy community if they are to survive

---

[1] The research project outlined in this paper represents collaborative work with colleagues Eva Baker, Leigh Burstein, Joan Herman, Daniel Koretz, and David Moody. It also draws heavily on other indicator development projects on which we have worked with Jeannie Oakes, Richard Shavelson, and Neil Carey.

[2] We include in the category of student coursework, measures of: The courses schools offer, patterns of course-taking by different types of students, the content of those courses, and the qualifications and experience of those teaching them.

as publicly supported endeavors. We also felt that a focus on student coursework was important because it helped address the oft-repeated admonition that any assessment of educational quality should be based on multiple indicators and should not depend solely on standardized test scores.

This paper examines the challenge of serving two masters by describing the major research and policy strands that shaped the SRA project, and then outlining how we attempted to meet those research and policy requirements in the study design. Because we have not completed our data collection, we cannot report on results, but in discussing our design, we do note some of the lessons about indicator development that we have learned along the way. The first section of the paper summarizes the major dimensions of curriculum that past research points to as important in the development of reliable and valid indicators. The second examines the key features of policy-relevant indicators; the third describes the current status of coursework indicators; and the final section outlines the SRA project design.

Reliable and Valid Coursework Indicators: Implications from Past Research

Past research on schooling strongly suggests that coursework indicators, or even the broader category of curriculum indicators, cannot be developed independently of a conceptual model of how the entire schooling system operates. Without such a model, single indicators can easily be taken out of context or misinterpreted (Guiton & Burstein, 1988; Shavelson, et al., 1987). Such a model should identify the major elements of the educational systems and illustrate the relationships among those elements. Clearly, it cannot specify relationships in either a strictly predictive or causal sense, but it can serve as a framework, showing logical linkages among components of the schooling system and correlational relationships supported by past research. In previous work on mathematics and science education indicators (Shavelson, et al., 1989), we drew upon others' research (e.g., Barr & Dreeben, 1983) to develop a model that includes three major components: (a) educational inputs—fiscal and other resources, teacher quality, and student background; (b) educational processes—school organization and context, curriculum, and instructional quality; and (c) educational outputs—student achievement, participation, and attitudes. Each of these elements is assumed to interact with each of the others. For example, curriculum quality shapes instructional activities and directly affects outputs such as achievement and attitudes. Similarly, curriculum quality is constrained by other factors such as teacher quality and the type and level of school resources.

In addition to the need to start with a comprehensive model of the educational system, past research suggests three factors important to developing reliable and valid curriculum indicators. First, curricula are highly differentiated: The content and treatment of mathematics courses, for example, differ substantially for various age and grade levels in elementary school, and for different ability levels in secondary schools (Oakes & Carey, 1989). The recent report of the National Research Council's Committee on Mathematics and Science Education Indicators recommends the creation of indicators based on four distinct curriculum "blocks"—classes for Grades K-5 and 6-8, 9-12 "academic" courses, and 9-12 "literacy" courses. Even within the academic set of courses, it may be useful to distinguish the curriculum intended for students who plan to major in mathematics and science at the university level from an academic, but non-science/mathematics major, curriculum (Murnane & Raizen, 1988). Indicators that attempt to describe the curriculum without being sensitive to these distinctions will obscure crucial attributes of the system, such as which students have access to what types of learning throughout their academic careers.

The second issue relates to the existence of several levels of curriculum within a school system. Simply put, this is the difference between the "intended" curriculum, as compared with the "implemented" curriculum (Murnane & Raizen, 1988). Actually, a continuum of levels exists, from the ideal conceptions of subject-matter and curriculum experts, on through various state and local policies, district curriculum guides, and teacher plans, to actual teacher-student interaction within the classroom (Oakes & Carey, 1989). The limitation of a curriculum indicator to any one of those levels poses a severe challenge to the validity of any inferences drawn from the indicator (Koretz, 1989). A robust indicator should incorporate information from a range of levels to provide as complete a description of what is taught as possible. Such an indicator should also allow a determination of the degree of "slippage" in curriculum implementation from one level to another (Oakes & Carey, 1989). Related to this issue is the question of the "achieved" curriculum, or what portion of the curriculum's objectives is actually acquired by students (Murnane & Raizen, 1988). However, this point may be more properly considered in the context of achievement indicators rather than curriculum measures, since achievement is influenced by a number of non-curricular factors such as student aptitudes (Koretz, 1989).

The issue of curriculum levels is particularly problematic for indicators of course offerings and student course-taking, since these indicators have traditionally focused on data collected at the school and district levels (e.g., graduation requirements, course offerings and enrollments by titles), and have rarely incorporated information about the course content or the nature of classroom interactions.

A third issue, not entirely distinct from the second, relates to which dimensions of the curriculum are actually being measured. A consensus seems to exist that content coverage is a key component, but that it alone is inadequate. Courses can be described in terms of the broad areas of content included in the syllabus, but such information needs to be supplemented with data on the depth and method of presentation. More detailed specifications of content coverage in "opportunity-to-learn" types of measures, where discrete topics and their relative emphasis within the curriculum are indicated, can be used. (For example, the Second International Mathematics Study incorporated a variety of "opportunity-to-learn" items in its teacher questionnaires; see McKnight, et al., 1987.)

Various other measures have been proposed, including some which draw on research in learning and instruction as well as curriculum. The National Research Council committee proposed collecting data on time spent on a lesson or the pages in a textbook allocated to various topics. Studies of students' conceptual and procedural difficulties in mastering the content of mathematics and science courses suggest areas to which a comprehensive curriculum indicator must be sensitive (Resnick, 1976, as cited in Oakes & Carey, 1989). For example, attention to the procedural and "metacognitive" aspects of mathematics (as opposed to the body of mathematical definitions and rules) is seen as a key element in mastering this discipline (Schoenfeld, 1985). Thus, the specific instructional goals and objectives of the curriculum form another important dimension, as does the sequencing of topics and the types of instructional activities used to realize the curriculum's goals. Another important dimension of the curriculum consists of the texts and materials used. This category includes textbooks, but also the constantly increasing body of supplementary materials, especially computer software developed for classroom use.

At various points, indicators of curriculum depth (as opposed to breadth of coverage) begin to overlap with other components of the educational system. For example, the instructional goals that teachers decide upon in the course of implementing curriculum in their classrooms could actually be considered a feature

of instructional strategies, as influenced by teacher qualifications and training. As in the case of student achievement and its relation to curriculum, this area may require treatment in a context other than that of curriculum indicator development (Carey, 1989).

Still, describing dimensions of the curriculum beyond simple content coverage is clearly necessary for developing robust indicators. Moreover, the additional information offers a way of addressing the first two issues described above. For instance, slippage across curriculum levels can be documented by information gathered on actual course objectives and types of instructional activities. As one example, where a state has mandated that more emphasis be placed on mathematics in the high school curriculum, a review of course objectives at the school or classroom levels can help determine whether additional mathematics courses are really contributing to students' mathematics education or whether they simply represent existing parts of the curriculum (e.g., lower-level mathematics or vocational courses) packaged in another format.

To summarize these issues and their implications for coursework indicators, we can point to three important requirements. First, sufficient information must be collected about courses and the students who enroll in them, so that courses can be characterized according to that part of the differentiated curriculum to which they belong. Second, an indicator must take into account different levels of the curriculum in order to present a valid description of its actual classroom-level manifestations. Third (and this relates to the second point), indicators of coursework need to incorporate measures that describe the various dimensions of what actually occurs within a given course. This is necessary to expand our understanding of what course titles actually signify, and as a way of addressing whether courses specified at one level of the system are the same as those implemented at another.

In addition to these requirements, it is also necessary to use curriculum indicators in conjunction with indicators measuring other aspects of the educational system. We know that the development and implementation of curriculum are constrained by school resources, student characteristics, and so on. For example, a curriculum indicator such as a measure of course-taking might show an increase in the proportion of students taking particular academic courses. Before we could interpret such a change as consistent with a policy directive to widen access to academic course-taking, we would have to check not just the content of those courses to make certain that they had not become less academically rigorous (and teacher assignment indicators to ascertain whether they were being taught by qualified teachers), but also trends in student demographics to make certain that a change in the nature of a school's or district's enrollment composition did not explain the change in course-taking behavior. In essence, indicators of other parts of the educational system are vital as checks on the construct validity of various curriculum indicators (Koretz, 1989).

Policy-Relevant Indicators: What Policymakers Need to Know

The first step in developing policy-relevant indicators is to recognize that once indicator data are viewed as pertinent to the policy community, they are no longer simply technical information. They will be interpreted and used in a political environment fueled by competing values and interests. For example, educational indicators can assist in identifying problems that, once recognized, become candidates for government action. Those wishing to expand the role of government may use such data to advance their position, while those espousing a more limited role may seek to discredit the data. Public support for education can be affected by

indicator data, though the direction might not be entirely predictable—for example, evidence of improved performance could result either in increased public support or in greater complacency. Indicator data can also lead policymakers and the public to focus on one aspect of a policy problem to the exclusion of others (MacRae, 1985). Consequently, those involved in developing policy-relevant indicators should recognize that their data may become politicized. Although they need to guard their own independence and neutrality, indicator designers can influence how information is likely to be used by taking features of the policy system into consideration from the beginning of the indicator development process. Understanding the characteristics of different policy audiences and of the broader policy context—how authority is distributed, what interests are represented, what issues are currently or likely to be on the policy agenda—will increase the likelihood that the data generated will be used appropriately.

A second step in developing indicators useful to the policy community and its constituents is to identify the generic information needs that can be addressed with indicator data and determine how those needs apply specifically to information on student coursework. Educational indicator data may be put to any of the following generic policy uses: (a) identifying and defining policy problems, (b) assessing the effects of existing policies, and (c) holding schools accountable.

By describing the current status of schooling and comparing it with that of earlier times or different places, indicator data can help in defining problems amenable to policy action. If a variety of data are collected within the framework of a comprehensive model, such information can also assist in identifying possible solutions through an analysis of trends in different indicators and the likely relationships among them.

Indicator data do not afford the level of detail and rigor needed for careful evaluations of individual policies or programs (MacRae, 1985; Shavelson, et al., 1989). However, such data can provide a partial basis for ascertaining the independent effects of discrete policy interventions—particularly, whether observed changes may be due to broader contextual factors rather than a particular policy. Indicator data can also suggest whether trends in the status of key educational indicators are consistent with what policymakers hope to achieve with particular types of policy. To serve that purpose, indicators need to be designed that capture the range of effects that particular categories of policy (e.g., those covering curriculum, teachers, etc.) are likely to produce: Indicator data cannot measure the effects of single policies, but it can signal larger changes that may result from a broader constellation of policy initiatives.

When used for accountability purposes, indicator data should meet several strict criteria (e.g., measure the central features of schooling, allow for fair comparisons, focus on the appropriate level of accountability—OERI State Accountability Study Group). Although no indicator systems yet meet these standards completely, the recent emphasis on improving indicators and developing better ones suggests that they may be used more appropriately for accountability purposes in the future.[3]

---

[3] However, as states attach more importance to indicators (e.g., by using the data to reward or punish schools), another threat to their validity and overall quality arises. The more important and salient an indicator is, the greater the likelihood that even a well-designed measure will be corrupted. An indicator becomes corrupted if those in the educational system "change their behavior in response to the indicator in a way that changes [its] meaning" (Koretz, in press). Common examples of indicators manipulated in ways that change their meaning include "teaching to the test" and altering course titles to conform to new policy regulations without changing the content.

To determine how these generic information needs applied to the development of coursework indicators, an initial phase of the SRA project relied on two approaches. In order to ensure that the indicators we developed captured the range of possible policy effects resulting from recent state educational reforms, we first needed to assess what had happened in local districts and schools as a result of those reforms, and what the implications were for indicator development. To do that, we examined the implementation and short-term effects of increased graduation requirements in five states—Arizona, California, Florida, Georgia, and Pennsylvania. We analyzed field interview data collected in the five state capitals, 19 local districts, and 30 high schools by researchers at RAND, Rutgers University, and the University of Wisconsin-Madison.[4]

Data from the five sample states suggested that, at a general level, policymakers' expectations in raising graduation requirements were quite similar across states. They wanted to improve student performance through more rigorous coursework, and to create uniform opportunities for academic coursework across different types of local districts and schools. At this level of generality, indicator data can measure whether changes in course offerings and student course-taking patterns are consistent with those expectations. However, such data cannot be used to ascertain whether policy shifts and changes in local practice are causally linked or even to provide very good explanations for why changes occurred in the way they did.

The analysis of local-level data suggested five major implications for our subsequent indicator development work. First, in developing coursework categories that can capture major changes in coursework policy, courses for both low- and high-achieving students need to be examined. The experience of the five sample states suggest that most new courses or additional sections were added at the lower end of a required subject. However, changes in college entrance requirements have also meant that coursework policies have affected a broader range of students than would have been the case with just the state mandates. The policy changes in the five sample states and the information from other states that have increased coursework requirements suggest that the subject areas with the greatest pay-off for developing policy-sensitive indicators are likely to be mathematics, science, and social studies.

Second, an indicator validation effort is needed that measures actual course content in a variety of ways (e.g., through teacher surveys, information about the text and coverage within it, sample lesson plans and assignments, etc.). This strategy is necessary because we found sufficient evidence to suggest that one response to state mandates has been to change course titles without significantly changing content; another has been to stratify even further courses within the same subject area and for schools to be quite explicit about it. Third, because we obtained conflicting reports in the five states about the effect of increased course requirements on academic stratification and on the high school drop-out rate, it is

---

[4] Across the five states and local districts, over 600 interviews were conducted. At the state level, approximately 150 people were interviewed during Spring 1986. These included: governors' education aides, state legislators and their staffs, state board of education members, state department of education officials, and interest group representatives.

Local district and school-level data were collected in February-March and May-June 1987. Interviews were conducted with local superintendents, school board members, district curriculum supervisors, teacher union leaders, principals, high school counselors, and 134 high school teachers.

These data are the same as those analyzed by William Clune (1989). Our findings about the implementation and effects of the high school graduation requirements are reported in McDonnell (1988), and are consistent with Clune's results.

important that indicator development be based on student samples before and after the state policy changes.

Fourth, data from the five states suggest that teaching out-of-field, as it relates to changes in coursework requirements, occurs within a very narrow band: Physical education and vocational education teachers have been moved into lower-level mathematics and science courses (although some out-of-field teaching is occurring in English and social studies). Therefore, if resources are limited, developing different measures of teacher mis-assignment could most profitably focus on teachers in those few subject areas. Finally, data from the five states suggest that reductions in course offerings, as a result of increases in other subjects, have also occurred within a fairly narrow band: The greatest reductions seem to have come in vocational education, social studies, and the arts and music. Therefore, if coursework indicators are to concentrate in the areas of greatest policy change, they should focus on the two or three areas most likely to have increased offerings and enrollments, and on the three listed above as those most likely to have been reduced or eliminated.

This first task provided guidance about the range of implementation patterns and effects associated with recent coursework policies. We also needed to know what policymakers themselves considered to be their major information needs in this area. Therefore, we undertook a second task which surveyed policymakers and their staffs about the types of coursework indicators that would be most useful to them and to their constituents. We conducted a focus group session with 20 governors' education aides and telephone interviews with 10 staff from national associations representing state policymakers (e.g., the National Conference of State Legislatures, the Council of Chief State School Officers) and 10 state and local policymakers and their staffs.[5]

Together, these three groups identified the following as the most pressing information needs in the area of student coursework policy:

1. Current data, based on simple enrollment counts, are inadequate. Policymakers are also interested in the content of those courses.

2. Respondents also expressed concern about the effects of new course policies on low-achieving students, but have little or no information on that issue.

3. Respondents also wanted to know about the unintended consequences of reform such as reduced curricular offerings or an increased drop-out rate.

4. Policymakers were interested in information about curricular opportunity costs—that is, if particular courses were required, what would that mean in terms of what students could no longer study?

5. Data that permit within-state comparisons were seen as more useful and important than across-state data.

Although the policy uses to which indicator data might be put vary considerably from strict research applications, our examination of the policy effects that coursework indicators need to capture and of policymakers' information needs

[5] The results of this survey are reported in Catterall, 1988.

suggests that considerable overlap exists between research and policy requirements. Both sets of requirements lead indicator designers to take into account: (a) the broader schooling context in which curriculum is delivered to students, (b) how it differs in content and treatment for different students, and (c) questions of depth as well as breadth of coverage. In addition to differences in their potential applications, one other distinction—though largely one of emphasis—between research and policy requirements is important to note. In constructing policy-relevant indicators, questions of feasibility are more salient than they might be for indicators designed primarily to serve research purposes. Indicators must not only be reliable, valid, and useful, they also must be able to be implemented within strict cost limits, not strain current levels of state and local expertise in data collection, analysis, and use, and create only a limited respondent burden on schools, teachers, and students (OERI State Accountability Study Group, 1988).

The Current Status of Coursework Indicators

Given what research and policy suggest as criteria for "good" coursework indicators, the next question is: How well do current indicators measure up to these standards, and what do they suggest for the next generation of indicators?

Among the various nationally representative databases in education, few include coursework indicators that satisfy the requirements outlined above, although a few now being developed are promising. Among the older studies, the National Assessment of Educational Progress (NAEP) in its several iterations since the 1960s includes some questions relating to course-taking and, to a lesser extent, types of instructional activity within courses. For example, students are asked about their experiences in science classes (e.g., whether they have ever performed an experiment). However, these responses do not relate to the particular courses that students have taken. On the other hand, principals are asked to describe patterns of course-taking within their schools and to estimate instructional time spent on various subjects.

Various longitudinal studies also include curriculum-related questions, but at about the same level of generality. The National Longitudinal Survey (NLS) of the class of 1972 and the High School and Beyond (HSB) study of the class of 1980 each included questions on course offerings and enrollments in their principal questionnaires. HSB also requested that students report on their course-taking, in terms of the number of classes taken within seven selected subject areas (mathematics, English, social studies, science, French, German, and Spanish), and whether they had taken any of six specified courses: second-year algebra, geometry, trigonometry, calculus, physics, and chemistry.

Although the usefulness of the course-taking data is limited by the lack of specific information about what was included in the courses, the reliability of the student-reported data has been established by studies done on the HSB and the course information collected on college-bound students who take the ACT. Both reports concluded that the information provided by students gives a fairly valid measure of student course-taking, although the validity varies among subject areas. The least amount of inaccuracy is found in subjects with fewer students (e.g. foreign languages and advanced science courses such as physics); and the most in those that are generally required of all students, but which are also prone to categorization difficulties, such as elective social studies and English classes (Fetters, et al., 1984; Valiga, 1987).

More recent studies such as the National Education Longitudinal Study (NELS) of 1988 are attempting to provide richer descriptions of curriculum at the

school and class levels. For example, the NELS asks teachers of each student in the survey to report on the topics covered in their classes and the emphasis placed on each topic, and about the materials and types of instructional activities used. Information is also requested about the ability level of the class, an item which can be employed in relating course content to particular "blocks" or tracks of the curriculum.

Although limited to the mathematics and science curriculum, the 1985 National Survey of Science and Math Education (NSSME) and the International Association for the Evaluation of Educational Achievement (IEA) studies of mathematics and science (conducted in 1981-82 and 1983-84/1986, respectively) provide similar types of information about coursework. The NSSME asked teachers about curriculum objectives, use of instructional time in a selected class, types of activities, and texts and materials used.[6] However, no information was requested about topics covered in the course, beyond the title of the class. The IEA gives somewhat more information, in that it asked teachers to report on their goals for the surveyed class, the topics covered, amount of time and emphasis allotted to each topic, and instructional approaches used. The IEA survey also included "opportunity-to-learn" measures, such as the teacher's assessment of whether their instruction was sufficient to allow students to answer test items correctly, and their estimate of their classes' probable success rate for each item. However, the usefulness of the IEA as an indicator is limited by problems of representativeness: Only about 50% of sampled school districts participated in the mathematics study; of the schools within districts sampled for the science study, 50% at the ninth-grade level responded, and only 36% at the fifth-grade level responded (Crosswhite, et al., 1985).

At the state level, a small number of states are beginning to collect information on actual implementation of the curriculum, but the majority appear to restrict themselves to promoting curriculum guidelines and limiting their curriculum-related data collection to course enrollment statistics and teacher assignment reports. The Council of Chief State School Officers (CCSSO) has found that in mathematics and science, 38 states have curriculum frameworks with a variety of purposes, including the establishment of a required curriculum or of mandated curriculum goals and objectives, the development of standardized tests, and the selection of texts and other materials. But although a number of state education agencies indicated that they were considering collecting data on school- and classroom-level curriculum implementation (including review of school curriculum outlines, teacher surveys, classroom observations, and "opportunity-to-learn" questionnaires), these were intended as "potential methods" of data collection rather than actual ones (Blank, 1988).

Although not intended specifically as a basis for indicator development, a notable exception to the typical state approach to coursework data is a Massachusetts study of course-taking (Massachusetts Department of Education, 1986) that attempted to assess the degree to which courses with identical titles varied in content and instruction across schools and across sections within schools. The study involved: (a) interviews with principals, counselors, department chairs, and teachers teaching algebra I and American history; (b) analysis of student transcripts; and (c) a detailed examination of course content in a small number of schools. Study findings on the extent of variation in content across courses with the same title adds

---

[6] Unlike the other nationally representative databases summarized in this section, the NSSME is based on data collected from surveys of principals and teachers, and does not include any student-level data.

credence to the notion that indicators of curriculum need to be based on data from several different levels of the educational system.

Typical indicators of course-taking at the state level include course enrollments and teacher assignments by course or subject area; in some cases these indicators now can be matched with information on teacher certification. Examples include: The Arkansas Department of Education's accreditation survey that asks each school to list course assignments and pupil loads for each teacher (the teacher lists can then be cross-referenced with a state-wide teacher credential file); and the California Basic Educational Data System (CBEDS) that asks for teachers' course assignments, class enrollments, and whether each class meets the University of California "a-f" requirements. The CBEDS also collects information on teacher qualifications. To the extent that the curriculum is affected by the type and extent of teacher training, both of these indicators could be said to relate to the "implemented" curriculum. However, this type of data obviously bears no relation to the content dimensions of curriculum—the breadth and depth of coverage that interest researchers and policymakers.

This brief review of currently available coursework indicators suggests that major gaps exist in measures that show how courses are differentiated across curriculum tracks—that is, the extent to which teacher qualifications, content, and instructional activities vary across courses aimed at different types of students. NELS '88 and NSSME, as newer nationally representative databases, have the capability to measure the extent of differentiation across different types of schools and classrooms. However, that capability does not extend to state-level indicator data, and since most curriculum policies emanate from states and local districts, data permitting that level of disaggregation are critical. Because the nationally representative databases are not intended to be linked to specific state or district policy contexts, we have virtually no existing indicators that measure the intended verses the implemented curriculum. Consequently, if we are to have indicators that are truly policy-relevant, design efforts will also need to be focused in that area. Existing indicators in the nationally representative databases are making great strides in measuring the breadth of the curriculum; depth has been a much more difficult dimension to measure, but improvements also are being made there. The challenge will be to translate those measures into ones that states and local districts might use, given the feasibility constraints discussed in the previous section. Finally, the notion that coursework indicators should be able to be linked conceptually to ones measuring other components of the educational system is now being reflected not just in the major national studies, but also in state-level indicators (as evidenced by efforts to link teacher assignment and certification data, and the school-level performance reports that states such as California and Illinois are now issuing). However, considerably more work will be needed in this area if research and policy requirements are to be effectively joined.

The SRA Project Design

Given the direction suggested from past research and the information needs identified by policymakers, our task in the SRA project was to design an indicator development effort that met the following criteria:

1. Concentrated on improving the reliability and validity of coursework indicators by more precisely distinguishing among blocks or tracks within the curriculum; providing a basis for comparing the implemented curriculum with expert standards or with policy objectives; and refining existing measures of the breadth and depth of content coverage.

2. Accommodated the information needs of policymakers with indicators that could capture, at least at a general level, the effects of major coursework policy initiatives.

3. Focused on course offerings and course-taking patterns, but also was sensitive to potential links with other indicators.

4. Made reasonable progress within the project's limited time frame (two years) and budget ($300,000).

These criteria led us to narrow our task to developing indicators of student course-taking that could be implemented by state governments as part of their existing indicator systems. In a sense, we would be developing a template that states could then field-test, and adapt to their own policy concerns, information needs, and data collection procedures. Such a focus meant that we would have to concentrate on measures for which data could be efficiently collected through surveys of school administrators, teachers, and students.

In designing a set of coursework indicators that state governments could adapt to their own data collection systems, we decided, where appropriate, to draw upon existing measures from sources such as SEA and NELS '88. However, because many non-cognitive items, typically used in routine indicator systems, have not been tested for their validity (Koretz, 1989), we also decided that a large part of our effort should consist of a validation study. Therefore, in addition to focusing on survey instruments of the type likely to be used by states, we also decided to undertake several benchmarking procedures—namely, by using interviews with school and district-level personnel, course materials, and student transcripts to verify data obtained from the surveys. Because the in-depth interviews and course material review provides information on coursework that is much closer to the actual content of instruction than are most routine indicators, they constitute criterion-related evidence of the validity of the survey data (Koretz, 1989). The transcript analysis will be an important source of historical data on how coursework patterns for different types of students have changed as compared with the pre-reform period, and thus provide a way of ascertaining whether the indicators we develop will still be valid if the nature of the curriculum were to change significantly. In sum, we decided that the major contribution of the SRA project would not be in developing entirely new indicators, but in refining existing ones, adapting them to the framework of state indicator systems, and above all, validating them through a number of benchmarking procedures.

Because of resource and time constraints, we chose to focus our indicator development effort on three course categories within mathematics—mathematics below algebra I (e.g., general math, consumer math, pre-algebra), algebra I, and algebra II—and two courses within social studies (American history and American government). These subjects were selected because they were among those most affected by state changes in high school graduation requirements; the specific course categories were chosen because our analysis of local responses to state curriculum policies suggested that the range of local effects could be captured largely with such a focus. Despite our limiting the development effort to only five course categories, we believed that the work could still serve as a template for future courses and subjects.

We chose to conduct the study in two different states, California and Georgia, because we wanted to control for the policy context in which indicators would be developed and used. Taking into account state policies will allow us to develop indicators that can be used in estimating the extent of curriculum "slippage" across levels of the educational system. We chose two states for which we already

had information on recent policies and local responses to those policies as a means of compressing the indicator development process. California's indicator system is among the most well-developed state systems in the country, but its information on student course-taking is limited to school-level enrollment statistics collected by course title (although the enrollment data for selected courses is disaggregated by student ethnicity). Because California has engaged in a major effort to upgrade its state-developed curriculum frameworks, it is particularly important that new indicators measure the extent to which that content is reflected in the school and classroom-level curriculum. Georgia is currently in the process of developing a more comprehensive state indicator system, and has appointed a task force to design a new course categorization system. Our study, then, is very timely, and should help in addressing a practical question that state officials have asked: "Can Georgia use a single course number for a course such as algebra I, or will we need multiple ones to distinguish among very different levels and content?" The Georgia system, which has three different diplomas (general, college preparatory, and vocational) each with different coursework requirements, also affords another basis for measuring curricular differentiation.

Within each state, we are using five high schools (Grades 9-12) as data sources for our validation study. Across the two states, four urban, three suburban, and three rural schools will be used. These schools in no way constitute a representative sample of high schools in California or Georgia. Not only did resource constraints limit us to such a small number, but the extent of data collection required in each school meant that for every school which agreed to participate, several others were contacted and refused.[7] However, in addition to differences in their location, we have also tried to use schools that vary in differentiation.[8]

The SRA project expects to end up with a set of coursework indicators that would allow policymakers to answer the following kinds of questions:

1. How much variation is there within individual schools and across different schools in the content of courses such as algebra I or American history?

2. How does this variation in content affect the learning opportunities of different kinds of students?

3. To what extent is the course content suggested (or mandated) by the state reflected in individual schools and classrooms?

4. What is the match between teacher qualifications and their course assignments, and in what courses is mis-assignment most prevalent?

---

[7] Despite the problems that we have experienced in gaining access to high schools, we made a decision not to reduce the scope of our data collection (e.g., by eliminating the transcript analysis or limiting the number of students surveyed). Had we done that, access would not have been a problem, but the quality of our validation effort would have been severely compromised. We also felt that since state governments would be the agencies most likely to field-test our indicators in the future, their authority to mandate such data collection would mean that the indicators would eventually be field-tested on an entire population of high schools or a representative sample of them.

[8] Of the six schools in which we have already collected data, four have a majority Anglo enrollment (55%-65%); one is majority Hispanic, and the other has an enrollment almost equally divided among Anglos, Blacks, and Hispanics. The schools vary in size from 332 to 2000 students; and the proportion of students attending four-year colleges ranges from 10% to 30%.

These indicators will be grouped according to the data collection procedures they require. For each set of indicators associated with a particular data collection strategy—course enrollment data from school rosters, teacher surveys, student surveys, and the benchmarking procedures of transcript and course materials analyses—we will assess how reliably and validly it measures course content and teacher assignment, and how feasible it is to collect and use. We expect that these different data sources represent a continuum, and that as we move from gross enrollment statistics to a course materials analysis, reliability and validity increase, but collection and use become significantly less feasible. However, by presenting our results as a comparative assessment of each set of indicators and associated data source, policymakers will be able to compare the marginal gain in preciseness of information with the trade-off in cost and burden.

Since this paper is really a report of research-in-progress, we conclude with a brief discussion of our five approaches to indicator validation and data collection: the kind of items included in each, the function it serves in the indicator development effort, and some problems or issues each has raised thus far.

Teacher Surveys. Because we assumed that these surveys would need to be administered as part of routine state data collection, we designed them to take teachers about 30 minutes to complete. In every school, all teachers who taught any mathematics or social studies course in the 1987-88 academic year are being surveyed. They are first asked questions about their educational background (e.g., number of mathematics or social studies courses, amount of subject-matter in-service over the past three years) and experience. They are then asked to give a period-by-period description of the classes they teach (including those outside mathematics and social studies as a means for understanding teacher assignment patterns), and to indicate whether and in what ways any of these courses may have been affected by recent changes in state graduation requirements or other state policies.

Those teachers teaching any of the five courses under study are then asked to complete a separate survey (still included in the 30-minute time limit) for each different section of the course that they teach. Teachers are asked about textbook and other materials, topic coverage;[9] the number of periods devoted to each topic, and whether it was taught as new content, reviewed and extended, reviewed only, assumed as prerequisite knowledge, or not taught and not assumed as student knowledge (the IEA strategy for ascertaining depth of coverage). Respondents are also asked about their instructional strategies (an adaptation of NAEP, IEA, and NELS '88 items), their goals for the course, the types of assignments and exams they gave, their distribution of grades, student preparation, and level of student performance, given that preparation.

Our very preliminary analysis of the teacher surveys that have been collected thus far suggests that some of the items that worked very well in the early 1980s as a means of distinguishing among different types of courses may have, in a sense, been corrupted by the reform rhetoric of the last few years. For example, there seems to be little variation among mathematics teachers in the emphasis they report giving to different curricular goals (e.g, developing an attitude of inquiry verses performing computations with speed and accuracy). However, within the

---

[9] In mathematics, the topics included on the survey (15 for algebra II and 23 for the other mathematics courses) are similar to those used in the IEA. For American history and government, we selected about 15 topics for each that included historical events, political institutions, and concepts (e.g., the potential conflict between liberty and equality). In choosing these, we relied on curriculum frameworks such as the new ones in California and consultations with several historians and political scientists.

teacher surveys themselves, we have some means of checking the validity of these responses. For example, even though a large proportion of teachers are reporting an emphasis on such "higher order" goals as understanding the nature of proof and the logical structure of mathematics, these same teachers do not report using instructional strategies consistent with those curricular goals. Our preliminary review of the data also suggests that some measure, not typically included in routine indicator systems, may be important in distinguishing among blocks or tracks of the curriculum. One example of such a factor is the distribution of grades, which seems to vary with a course's ability level.

Student surveys. These surveys were conceived as the type of questionnaire that states could administer in conjunction with their standardized achievement tests. Consequently, the student surveys are even shorter than those administered to teachers—approximately 10 minutes for 10th graders and 15-20 minutes for 12th graders. These were administered to all 10th and 12th graders in attendance on one particular day (yielding completed surveys from 66%-95% of all current 10th and 12th graders, depending on the school). These surveys are designed in such a way that they can be linked to individual teachers. In addition to including items about the student's background and future educational plans, the surveys repeat the instructional strategy questions asked of teachers. In this way we will be able to compare the reliability of these two data sources.

Transcript data. In each school, 75 transcripts were randomly sampled from those students who were ninth graders in 1982 (1983 in Georgia), 1986, and 1988 (for a total of 225 transcripts per school). We used the ninth-grade class as the sampling base to ensure that we included students who may not have completed high school. These three class years were selected because those who graduated in 1986 in California and 1987 in Georgia represent the last class to progress through high school before state-mandated increases in course requirements were applicable; the class of 1989 is one of the first classes under the new requirements and allows us to examine course-taking by a class that took American history the prior year and is currently enrolled in government (some students will have also taken algebra II the prior year); the class of 1991 provides an opportunity to examine the previous year's course-taking in lower-level math and algebra I.

Each transcript is being coded to include student background (gender, ethnicity, birthdate, GPA, standardized test scores, number of absences). For each course (in mathematics, social studies, English, science, foreign language, vocational education, and fine arts), the following information is coded: (a) whether it is remedial, basic/regular, college prep, honors, advanced placement, applied, or "everyperson" (i.e., a course open to all students, such as electives or required courses in schools with minimal ability grouping); (b) whether it is intended for a special population such as handicapped or limited-English-proficiency students when the course was taken; (c) the grade a student received; and (d) whether it was taken at the school under study or is transferred credit from another school. In categorizing courses, we did not rely on existing course categorization schemes (e.g., the one used by HSB), but rather devised our own by examining each course in our sample schools and creating categories that were meaningful across those schools. In most cases, our categories are quite similar to other coding schemes, but this exercise provided another way to validate the information we were collecting from other sources.

The transcript analysis will be key to our efforts to understand (a) how the curriculum is differentiated within a particular school and (b) the course-taking patterns associated with different types of students. The analysis also will provide a source of validation for what teachers tell us about how coursework has changed as a result of state and local policy.

Our major problem with the student transcript data, thus far, has been our inability to obtain a valid or reliable measure of students' socio-economic status. Since most transcripts do not contain information about parental occupation, that obvious measure was not available. The only measures we have is whether a student lives in a home with two parents, a single parent, or a guardian, and whether he or she is eligible for free or reduced-price lunch.[10]

In-depth interviews. In each school we interviewed the principal, the head counselor, and the chairs of the mathematics and social studies departments. We also interviewed the district-level person responsible for supervising the high school curriculum. These interviews typically lasted about one hour and were often followed up by additional telephone inquiries. The purpose of these interviews was to understand: (a) The type of students attending the school and whether it had changed recently, (b) the different levels of courses offered and whether this differentiation of the curriculum has the same meaning across different departments, (c) what criteria the school uses in assigning students to different courses and sections, (d) how decisions about teacher assignment are made, and (e) how recent state policies may have affected the school's course offerings and instructional practices. In the interviews with department chairs, we also asked them to describe in some detail the major differences among the five courses we are examining in terms of: (a) level of difficulty, (b) the types of students enrolled, (c) topics covered, (d) instructional materials and strategies, (e) course requirements, and (f) grading practices. In the interview with the district-level staff, we were particularly interested in district policies that were intended to influence the school-level curriculum, and how the sample school compared with others in the district in terms of its course offerings and student assignment policies.

These interviews have been critical in creating meaningful course categories for the transcript analysis, and will provide an important source of validation when we begin to analyze the survey data.

Course materials. This last data source has been the most problematical for us. We had originally hoped to collect sample assignments, as well as course syllabi and final exams. However, we realized that such an effort would be burdensome to teachers, and would be difficult for us to interpret validly (e.g., is the collected assignment really a typical assignment for the third week of the semester or is it a teacher's "best" or "most difficult" assignment? We would not be able to determine that even with much additional effort.) Consequently, we decided only to request a copy of each surveyed teacher's syllabus (asking how much was covered in last year's class) and their final examination. Even this scaled-down information has been difficult to obtain. Only about half the teachers in our sample have been able to provide both pieces of material because many do not retain syllabi and exams from one year to the next. We have put additional effort into this area, but it remains a problem. Nevertheless, we have found that even with limited course materials, this source is serving an important validation function for the teacher surveys (e.g., by comparing stated topic coverage and curricular goals with final exams).

Conclusion (or Actually the Lack of One)

This story of attempting to serve two master lacks an ending. Until we complete our data collection and analysis, we will not know whether we have been

---

[10] This later measure is not particularly reliable, especially in urban high schools where a large number of eligible students do not apply for reduced price or free lunch because of the preceived stigma.

successful in our efforts to join research requirements and policy needs. However, even if our final product fails to meet all our initial expectations, we believe that the process of developing reliable and valid coursework indicators that can be used in a statewide indicator system holds important lessons for future efforts to link research and policy information.

References

Barr, R., & Dreben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Blank, R. K. (1988). *State data available on indicators of science/math education.* Draft paper. Washington, DC: Council of Chief State School Officers.

Carey, N. (1989). Instruction. In R.J. Shavelson, L.M. McDonnell, & J. Oakes (Eds.), *Indicators for monitoring mathematics and science education: A sourcebook.* Santa Monica, CA: The Rand Corporation.

Catterall, J. (1988). *An assessment of audience needs for school reform assessment information.* Los Angeles: UCLA Center for the Study of Evaluation.

Clune, W.H., White, P., & Patterson, J. (1989). *The implementation and effects of high school graduation requirements: First steps toward curricular reform.* New Brunswick, NJ: Center for Policy Research in Education.

Crosswhite, F.J., Dossey, J.A., Swafford, J.O., McKnight, C.C., & Cooney, T.J. (1985). *Second International Mathematics Study: Summary Report for the United States.* Washington, DC: National Center for Education Statistics.

de Neufville, J.I. (1975). *Social indicators and public policy: Interactive processes of design and application.* New York: Elsevier Scientific Publishing Company.

Fetters, W.B., Stowe, P.S., & Owings, J.A. (1984). *High school and beyond: Quality of responses of high school students to questionnaire items.* Washington, DC: National Center for Education Statistics.

Guiton, G., & Burstein, L. (1988). *States in the construction of educational indicator systems.* Draft paper. Los Angeles: UCLA Center for the Study of Evaluation.

Koretz, D. (1989). *A framework for evaluating and validating indicators of mathematics and science education.* Santa Monica, CA: The Rand Corporation.

MacRae, D., Jr. (1985). *Policy indicators.* Chapel Hill, NC: University of North Carolina Press.

Massachusetts Department of Education. (1986). *The high school experience in Massachusetts.* Boston, MA.

McDonnell, L. (1988). *Coursework policy in five states and its implications for indicator development.* Working paper. Santa Monica, CA: The Rand Corporation.

McKnight, C.C., Crosswhite, F.J., Dossey, J.A., Kifer, E., Swafford, S.O., Travers, K.J., & Cooney, T.J. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes Publishing.

Murnane, R.J., & Raizen, S.A. (Eds.). (1988). *Improving indicators of the quality of science and mathematics education in grades K-12.* Washington, DC: National Academy Press.

Oakes, J., & Carey, N. (1989). Curriculum. In R.J. Shavelson, L.M. McDonnell, & J. Oakes (Eds.), *Indicators for monitoring mathematics and science education: A sourcebook.* Santa Monica, CA: The Rand Corporation.

OERI State Accountability Study Group. (1988). *Creating responsible and responsive accountability systems.* Washington, DC: U.S. Department of Education.

Resnick, L.B. (1976). Task analysis in instructional design: Some cases from mathematics. In D. Klahr (Ed.), *Cognition and instruction.* New York: Lawrence Erlbaum.

Schoenfeld, A.H. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press, Inc.

Shavelson, R.J., McDonnell, L.M., & Oakes, J. (Eds.). (1989). *Indicators from Monitoring Mathematics and Science Education: A Sourcebook.* Santa Monica, CA: The Rand Corporation.

Shavelson, R., McDonnell, L., Oakes, J., Carey, N., & Picus, L. (1987). *Indicator systems for monitoring mathematics and science education.* Santa Monica, CA: The Rand Corporation

Valiga, M.J. (1987). *The accuracy of self-reported high school course and grade information.* Iowa City, IA: The American College Testing Program.